Bounds on the Capacity of Random Insertion and Deletion-Additive Noise Channels

Mojtaba Rahmati, *Student Member, IEEE*, and Tolga M. Duman, *Fellow, IEEE*

Abstract—We develop several analytical lower bounds on the capacity of binary insertion and deletion channels by considering independent uniformly distributed (i.u.d.) inputs and computing lower bounds on the mutual information between the input and output sequences. For the deletion channel, we consider two different models: i.i.d. deletion–substitution channel and i.i.d. deletion channel with additive white Gaussian noise (AWGN). These two models are considered to incorporate effects of the channel noise along with the synchronization errors. For the insertion channel case, we consider Gallager's model in which the transmitted bits are replaced with two random bits and uniform over the four possibilities independently of any other insertion events. The general approach taken is similar in all cases, however the specific computations differ. Furthermore, the approach yields a useful lower bound on the capacity for a wide range of deletion probabilities of the deletion channels, while it provides a beneficial bound only for small insertion probabilities (less than 0.25) of the insertion model adopted. We emphasize the importance of these results by noting that: 1) our results are the first analytical bounds on the capacity of deletion-AWGN channels, 2) the results developed are the best available analytical lower bounds on the deletion–substitution case, 3) for the Gallager insertion channel model, the new lower bound improves the existing results for small insertion probabilities.

Index Terms—Achievable rates, channel capacity, insertion/deletion channels, synchronization.

I. INTRODUCTION

I N modeling digital communication systems, we often assume that the transmitter and receiver are completely synchronized; however, achieving a perfect time-alignment between the transmitter and receiver clocks is not possible in all communication systems and synchronization errors are unavoidable. A useful model for synchronization errors assumes that the number of received bits may be more or less than the number of transmitted bits. In other words, insertion/deletion channels may be used as appropriate models for communication channels that suffer from synchronization errors. Due

to the memory introduced by the synchronization errors, an information theoretic study of these channels proves to be very challenging. For instance, even for seemingly simple models such as an i.i.d. deletion channel, an exact calculation of the capacity is not possible and only upper/lower bounds (which are often loose) are available.

In this paper, we compute analytical lower bounds on the capacity of the i.i.d. deletion channel with substitution errors and in the presence of AWGN, and i.i.d. random insertion channel, by lower bounding the mutual information rate between the transmitted and received sequences for i.u.d. inputs. We particularly focus on the small insertion/deletion probabilities with the premise that such small values are more practical from an application point of view, where every bit is independently deleted with probability p_d or replaced with two randomly chosen bits with probability p_i, while neither the transmitter nor the receiver have any information about the positions of deletions and insertions, and undeleted bits are flipped with probability p_e and bits are received in the correct order. By a deletion–substitution channel, we refer to an insertion/deletion channel with $p_i = 0$; by a deletion-AWGN channel we refer to an insertion/deletion channel with $p_i = p_e = 0$ (deletion-only channel) in which undeleted bits are received in the presence of AWGN, that can be modeled by a combination of a deletion-only channel with a binary input AWGN (BI-AWGN) channel such that every bit first goes through a deletion-only channel and then through a BI-AWGN channel. Finally, by a random insertion channel we refer to an insertion/deletion channel with $p_d = p_e = 0$.

A. Review of Existing Results

Dobrushin [1] proved under very general conditions that for a memoryless channel with synchronization errors, Shannon's theorem on transmission rates applies and the information and transmission capacities are equal. The proof hinges on showing that information stability holds for the insertion/deletion channels and, as a result [2], capacity per bit of an i.i.d. insertion/deletion channel can be obtained by $\lim_{N \to \infty} \max_{P(X)} \frac{1}{N} I(X; Y)$, where X and Y are the transmitted and received sequences, respectively, and N is the length of the transmitted sequence. On the other hand, there is no single-letter or finite-letter formulation which may be amenable for the capacity computation, and no results are available providing the exact value of the limit.

Gallager [3] considered the use of convolutional codes over channels with synchronization errors, and derived an expression which represents an achievable rate for channels with insertion, deletion, and substitution errors (whose model is specified earlier). The approach is to consider the transmission of i.u.d. binary information sequences by convolutional coding and

modulo-2 addition of a pseudorandom binary sequence (which could be considered as a watermark used for synchronization purposes), and computation of a rate that guarantees a successful decoding by sequential decoding. The achievable rate, or the capacity lower bound, is given by the expression

$$C \geq 1 + p_d \log p_d + p_i \log p_i + p_c \log p_c + p_s \log p_s \quad (1)$$

where C is the channel capacity, $p_c = (1 - p_d - p_i)(1 - p_e)$ is the probability of correct reception, and $p_s = (1 - p_d - p_i)p_e$ is the probability that a flipped version of the transmitted bit is received. The logarithm is taken base 2 resulting in transmission rates in bits/channel use. By substituting $p_i = 0$ in (1), for $p_d \leq 0.5$, a lower bound on the capacity of the deletion–substitution channel C_{ds} can be obtained as

$$C_{ds} \geq 1 - H_b(p_d) - (1 - p_d)H_b(p_e) \quad (2)$$

where $H_b(p_d) = -p_d \log p_d - (1 - p_d) \log(1 - p_d)$ is the binary entropy function. It is interesting to note that for $p_d = p_e = 0$ ($p_i = p_e = 0$) and $p_i \leq 0.5$ ($p_d \leq 0.5$), a lower bound on the capacity of the random insertion channel (deletion-only channel) with insertion (deletion) probability of p_i (p_d) is equal to the capacity of a binary symmetric channel (BSC) with a substitution error probability of p_i (p_d).

In [4] and [5], authors argue that, since the deletion channel has memory, optimal codebooks for use over deletion channels should have memory. Therefore, in [4]–[7], achievable rates are computed by using a random codebook of rate R with $2^{n \cdot R}$ codewords of length n, while each codeword is generated independently according to a symmetric first-order Markov process. Then, the generated codebook is used for the transmission over the i.i.d. deletion channel. In the receiver, different decoding algorithms are proposed, e.g., in [4], if the number of codewords in the codebook that contain the received sequence as a subsequence is only one, the transmission is successful, otherwise an error is declared. The proposed decoding algorithms result in an upper bound for the incorrect decoding probability. Finally, the maximum value of R that results in a successful decoding as $n \rightarrow \infty$ is an achievable rate, hence a lower bound on the transmission capacity of the deletion channel. The lower bound (1), for $p_i = p_e = 0$, is also proved in [4] using a different approach compared to the one taken by Gallager [3], where the authors computed achievable rates by choosing codewords randomly, independently, and uniformly among all possible codewords of a certain length.

In [8], a lower bound on the capacity of the deletion channel is directly obtained by lower bounding the information capacity $\lim_{N \rightarrow \infty} \frac{1}{N} \max_{P(\boldsymbol{X})} I(\boldsymbol{X}; \boldsymbol{Y})$. In [8], input sequences are considered as alternating blocks of zeros and ones (runs), where the length of the runs L are i.i.d. random variables following a particular distribution over positive integers with a finite expectation and finite entropy ($E(L), H(L) < \infty$ where $E(\cdot)$ and $H(\cdot)$ denote the expected value and entropy, respectively).

In [9] and [10], Monte Carlo methods are used for computing lower bounds on the capacity of the insertion/deletion channels based on reduced-state techniques. In [9], the input process is assumed to be a stationary Markov process and lower bounds

on the capacity of the deletion and insertion channels are obtained via Monte Carlo simulations considering both the first and second-order Markov processes as input. In [10], information rates for i.u.d. input sequences are computed for several channel models using a similar Monte Carlo approach where in addition to the insertions/deletions, effects of intersymbol interference (ISI) and AWGN are also investigated.

There are several papers deriving upper bounds on the capacity of the insertion/deletion channels as well. Fertonani and Duman in [11] present several novel upper bounds on the capacity of the i.i.d. deletion channel by providing the decoder (and possibly the encoder) with some genie-aided information about the deletion process resulting in auxiliary channels whose capacities are certainly upper bounds on the capacity of the i.i.d. deletion channel. By providing the decoder with appropriate side information, a memoryless channel is obtained in such a way that Blahut–Arimoto algorithm can be used for evaluating the capacity of the auxiliary channels (or, at least computing a provable upper bound on their capacities). They also prove that by subtracting some value from the derived upper bounds, lower bounds on the capacity can be derived. The intuition is that the subtracted information is more than extra information added by revealing certain aspects of the deletion process. A nontrivial upper bound on the deletion channel capacity is also obtained in [12] where a different genie-aided decoder is considered. Furthermore, Fertonani and Duman in [13] extend their work [11] to compute several upper and lower bounds on the capacity of channels with insertion, deletion, and substitution errors as well.

In two recent papers [14], [15], asymptotic capacity expressions for the binary i.i.d. deletion channel for small deletion probabilities are developed. In [15], the authors prove that $C_d \leq 1 - (1 - O(p_d))H_b(p_d)$ (where $O(.)$ represents the standard Landau (big-O) notation) which clearly shows that for small deletion probabilities, $1 - H_b(p_d)$ is a tight lower bound on the capacity of the deletion channel. In [14], an expansion of the capacity for small deletion probabilities is computed with several dominant terms in an explicit form. The interpretation of our result for i.i.d. deletion-only channel case is parallel to the one in [15].

B. Contributions of the Paper

In this paper, we focus on small insertion/deletion probabilities and derive analytical lower bounds on the capacity of the insertion/deletion channels by lower bounding the mutual information between i.u.d. input sequences and resulting output sequences. Since as shown in [1], for an insertion/deletion channel, the information and transmission capacities are equal justifying our approach in obtaining an achievable rate.

We note that our idea is somewhat similar to the idea of directly lower bounding the information capacity instead of lower bounding the transmission capacity as employed in [8]. However, there are fundamental differences in the main methodology as will become apparent later. For instance, our approach provides a procedure that can easily be employed for many different channel models with synchronization errors as such we are able to consider deletion–substitution, deletion-AWGN, and random insertion channels. Other differences include adopting a finite-length transmission which is proved to yield a lower

bound on the capacity after subtracting some appropriate term, and the complexity in computing the final expression numerically is much lower in many versions of our results.

Finally, we emphasize that by utilizing the new approach, we improve upon the obtained results in the existing literature in several different aspects. In particular, the contributions of the paper include:

1) development of a new approach for deriving achievable information rates for insertion/deletion channels,
2) the first analytical lower bound on the capacity of the deletion-AWGN channel,
3) tighter analytical lower bounds on the capacity of the deletion–substitution channel for all values of deletion and substitution probabilities compared to the existing analytical results,
4) tighter analytical lower bounds on the capacity of the random insertion channels for small values of insertion probabilities ($p_i < 0.25$) compared to the existing lower bounds,
5) very simple lower bounds on the capacity of several cases of insertion/deletion channels.

Regarding the final point, we note that by employing $p_e = 0$ in the results on the deletion–substitution channel, we arrive at lower bounds on the capacity of the deletion-only channel which are in agreement with the asymptotic results of [14], [15] in the sense of capturing the dominant terms in the capacity expansion. Our results, however, are provable lower bounds on the capacity, while the existing asymptotic results are not amenable for numerical calculation (as they contain big-O terms).

C. Notation

We denote a binary sequence of length n with K runs by $(b; n_1, n_2, \ldots, n_K)$, where $b \in \{0, 1\}$ denotes the first run type and $\sum_{k=1}^{K} n_k = n$. For example, the sequence 001111011000 can be represented as (0;2,4,1,2,3). We use four different ways to denote different sequences; $\boldsymbol{x}(b; n^x; K^x)$ represents every sequence belonging to the set of sequences of length n^x with K^x runs and by the first run of type b, $\boldsymbol{x}(b; n^x; K^x; l)$ represents a sequence $\boldsymbol{x}(b; n^x; K^x)$ which has l runs of length one ($l = \sum_{k=1}^{K^x} \delta(n_k^x - 1)$ with $\delta(.)$ denoting the Kronecker delta function), $\boldsymbol{x}(n^x)$ represents every sequence of length n^x, and $\boldsymbol{x}$ represents every possible sequence. The set of all input sequences is shown by $\mathcal{X}$, and the set of output sequences of the deletion-only, and random insertion channels are shown by $\mathcal{Y}^d$ and $\mathcal{Y}^i$, respectively. $\mathcal{Y}^d_{-a}$ and $\mathcal{Y}^i_{+c}$ denote the set of output sequences resulting from a deletions and c random insertions, respectively, and $\mathcal{Y}^d(\boldsymbol{x} - a)$ and $\mathcal{Y}^i(\boldsymbol{x} + c)$ denote the set of output sequences resulting from a deletions from and c random insertions into, the input sequence $\boldsymbol{x}$, respectively. We denote the deletion pattern of length d in a sequence of length n with K runs by $D(n; K; d) = (d_1, d_2, \ldots, d_K)$, where d_k denotes the number of deletions in the kth run and $\sum_{k=1}^{K} d_k = d$. The outputs resulting from a given deletion pattern $D(n; K; d) = (d_1, d_2, \ldots, d_K)$ (without any other error) are denoted by $D(n; K; d) * \boldsymbol{x}(n; K) = (n_1 - d_1, n_2 - d_2, \ldots, n_K - d_K)$. The set $\mathcal{D}_K^n(d)$ represents the set of all deletion patterns of length d of a sequence of length n and with K runs.

D. Organization of the Paper

In Section II, we introduce our general approach for lower bounding the mutual information of the input and output sequences for insertion/deletion channels. In Section III, we apply the introduced approach to the deletion–substitution and deletion-AWGN channels and present analytical lower bounds on their capacities, and compare the resulting expressions with earlier results. In Section IV, we provide lower bounds on the capacity of the random insertion channels and comment on our results with respect to the existing literature. In Section V, we compute the lower bounds for a number of insertion/deletion channels, and finally, we provide our conclusion in Section VI.

II. MAIN APPROACH

We rely on lower bounding the information capacity of memoryless channels with insertion or deletion errors directly as justified by [1], where it is shown that, for a memoryless channel with synchronization errors, the Shannon's theorem on transmission rates applies and the information and transmission capacities are equal, and thus every lower bound on the information capacity of an insertion/deletion channel is a lower bound on the transmission capacity of the channel. Our approach is different than most existing work on finding lower bounds on the capacity of the insertion/deletion channels where typically the transmission capacity is lower bounded using a certain codebook and particular decoding algorithms. The idea we employ is similar to the work in [8] which also considers the information capacity $\lim_{N \to \infty} \frac{1}{N} \max_{P(\boldsymbol{X})} I(\boldsymbol{X}; \boldsymbol{Y})$ and directly lower bounds it using a particular input distribution to arrive at an achievable rate result.

Our primary focus is on the small deletion and insertion probabilities. As also noted in [14], for such probabilities it is natural to consider binary i.u.d. input distribution. This is justified by noting that when $p_d = p_i = 0$, i.e., for a binary symmetric channel, the capacity is achieved with independent and symmetric binary inputs, and hence we expect that for small insertion/deletion probabilities, binary i.u.d. inputs are not far from the optimal input distribution.

Our methodology is to consider a finite length transmission of i.u.d. bits over the insertion/deletion channels, and to compute (more precisely, lower bound) the mutual information between the input and the resulting output sequences. As proved in [11] for a channel with deletion errors, such a finite length transmission in fact results in an upper bound on the mutual information supported by the insertion/deletion channels; however, as also shown in [11], if a suitable term is subtracted from the mutual information, a provable lower bound on the achievable rate, hence the channel capacity, results. The following theorem provides this result in a slightly generalized form compared to [11].

Theorem 1: For binary input channels with i.i.d. insertion or deletion errors, for any input distribution and any $n > 0$, the channel capacity C can be lower bounded by

$$C \geq \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y}) - \frac{1}{n} H(\boldsymbol{T}) \qquad (3)$$

where

$$H(\boldsymbol{T}) = -\sum_{j=0}^{n} \left[\binom{n}{j} p^j (1-p)^{n-j} \log \left(\binom{n}{j} p^j (1-p)^{n-j} \right) \right]$$

with the understanding that $p = p_d$ for the deletion channel case and $p = p_i$ in the insertion channel case, and n is the length of the input sequence $\boldsymbol{X}$.

Proof: This is a slight generalization of a result in [11] which shows that (3) is valid for the i.i.d. deletion channel. It is easy to see that [11], for any random process $\boldsymbol{T}^N$ and for any input distribution $P(\boldsymbol{X}^N)$, we have

$$C \geq \lim_{N \to \infty} \frac{1}{N} I(\boldsymbol{X}^N; \boldsymbol{Y}^N, \boldsymbol{T}^N) - \lim_{N \to \infty} \frac{1}{N} H(\boldsymbol{T}^N) \quad (4)$$

where C is the capacity of the channel, N is the length of the input sequence $\boldsymbol{X}^N$, and $N = Qn$, i.e., the input bits in both insertion and deletion channels are divided into Q blocks of length n ($\boldsymbol{X}^N = \{\boldsymbol{X}_j\}_{j=1}^Q$). We define the random process $\boldsymbol{T}^N$ in the following manner. For an i.i.d. insertion channel, $\boldsymbol{T}^{N,i}$ is formed as the sequence $\boldsymbol{T}^{N,i} = \{T_j^i\}_{j=1}^Q$ which denotes the number of insertions that occur in transmission of each block of length n. For a deletion channel, $\boldsymbol{T}^{N,d} = \{T_j^d\}_{j=1}^Q$ represents the number of deletions occurring in transmission of each block. Since insertions (deletions) for different blocks are independent, the random variables $T_j = T_j^i$ (T_j^d) for $j \in \{1, \ldots, Q\}$ are i.i.d., and transmission of different blocks are independent. Therefore, we can rewrite (4) as

$$C \geq \frac{1}{n} I(\boldsymbol{X}_j; \boldsymbol{Y}_j) - \frac{1}{n} H(\boldsymbol{T}_j)$$
$$= \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y}) - \frac{1}{n} H(\boldsymbol{T}). \quad (5)$$

Noting that the random variable denoting the number of deletions or insertions as a result of n bit transmission is binomial with parameters n and p_d (or, p_i) the result follows. ∎

Several comments on the specific calculations involved are in order. Theorem 1 shows that for any input distribution and any transmission length, (3) results in a lower bound on the capacity of the channel with deletion or insertion errors. Therefore, employing any lower bound on the mutual information rate $\frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y})$ in (3) also results in a lower bound on the capacity of the insertion/deletion channel. Due to the fact that obtaining the exact value of the mutual information rate for any n is infeasible, we first derive a lower bound on the mutual information rate for i.u.d. input sequences and then employ it in (3). Based on the formulation of the mutual information, obviously

$$I(\boldsymbol{X}; \boldsymbol{Y}) = H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X}) \quad (6)$$

thus by calculating the exact value of the output entropy or lower bounding it and obtaining the exact value of the conditional output entropy or upper bounding it, the mutual information is lower bounded. For the models adopted in this paper, we are able to obtain the exact value of the output sequence probability distribution when i.u.d. input sequences are used, hence the exact value of the output entropy (the differential output entropy for the deletion-AWGN channel) is available.

In deriving the conditional output entropies (the conditional differential entropy of the output sequence for the deletion-AWGN channel), we cannot obtain the exact probability of all the possible output sequences conditioned on a given input sequence. For deletion channels, we compute the probability of all possible deletion patterns for a given input sequence,
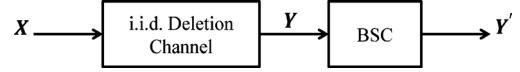

Fig. 1. Deletion–substitution channel as a cascade of an i.i.d. deletion channel and a BSC.

and treat the resulting sequences as if they are all distinct to find a provable upper bound on the conditional entropy term. Clearly, we are losing some tightness, as different deletion patterns may result in the same sequence at the channel output. For the random insertion channel, we calculate the conditional probability of the output sequences resulting from at most one insertion, and derive an upper bound on the part of the conditional output entropy expression that results from the output sequences with multiple insertions.

III. LOWER BOUNDS ON THE CAPACITY OF NOISY DELETION CHANNELS

As mentioned earlier, we consider two different variations of the binary deletion channel: i.i.d. deletion and substitution channel (deletion–substitution channel), and i.i.d. deletion channel in the presence of AWGN (deletion-AWGN channel). The results utilize the idea and approach of the previous section. We first give the results for the deletion–substitution channel, then for the deletion-AWGN channel. We note that the presented lower bounds can be also employed on the deletion-only channel if $p_e = 0$ (or $\sigma^2 = 0$ for the deletion-AWGN channel).

A. Deletion–Substitution Channel

In this section, we consider a binary deletion channel with substitution errors in which each bit is independently deleted with probability p_d, and transmitted bits are independently flipped with probability p_e. The receiver and the transmitter do not have any information about the position of deletions or the substitution errors. As shown in Fig. 1, this channel can be considered as a cascade of an i.i.d. deletion channel with a deletion probability p_d and output sequence $\boldsymbol{Y}$, and a BSC with a cross-over error probability p_e and output sequence $\boldsymbol{Y}'$. For such a channel model, the following lemma is a lower bound on the capacity.

Lemma 1: For any $n > 0$, the capacity of the i.i.d. deletion–substitution channel C_{ds}, with a substitution probability p_e and a deletion probability p_d, is lower bounded by

$$C_{ds} \geq 1 - p_d - H_b(p_d) - (1 - p_d) H_b(p_e)$$
$$+ \frac{1}{n} \sum_{j=1}^n W_j(n) \binom{n}{j} p_d^j (1 - p_d)^{n-j} \quad (7)$$

where

$$W_j(n) = \sum_{l=1}^{n-1} 2^{-l-1} (n - l + 3) \sum_{j'=1}^j \frac{\binom{l}{j'} \binom{n-l}{j-j'}}{\binom{n}{j}} \log \binom{l}{j'}$$
$$+ 2^{-n+1} \log \binom{n}{j} \quad (8)$$

and $H_b(p_d) = -p_d \log(p_d) - (1 - p_d) \log(1 - p_d)$. ∎

Before proving the lemma, we would like to emphasize that the only existing analytical lower bound on the capacity of deletion–substitution channels is derived in [3] ((2)). In comparing the lower bound in (2) with the lower bound in (7), we observe that the new lower bound improves the previous one by $\frac{1}{n} \sum_{j=1}^{n} W_j(n) \binom{n}{j} p_d^j (1 - p_d)^{n-j} - p_d$, which is guaranteed to be positive.

A simplified form of the lower bound for small values of deletion probability can also be presented. By invoking the inequalities $(1 - p)^m \geq [1 - mp + \binom{m}{2} p^2 - \binom{m}{3} p^3]$ and $(1 - p)^m \geq 1 - mp$, and ignoring some positive terms $(p_d^j (1 - p_d)^{n-j}$ for $j \geq 3)$, we can write

$$C_d \geq 1 - H_b(p_d) + p_d(W_1(n) - 1) + p_d^2 \frac{n-1}{2} (W_2(n) - 2W_1(n))$$
$$+ p_d^3 \binom{n-1}{2} (W_1(n) - W_2(n)) - p_d^4 \binom{n-1}{3} W_1(n).$$

By utilizing $p_e = 0$ in (7), we can obtain a lower bound on the capacity of the deletion-only channel as given in the following corollary.

Corollary 1: For any $n > 0$, the capacity of an i.i.d. deletion channel C_d, with a deletion probability of p_d is lower bounded by

$$C_d \geq 1 - p_d - H_b(p_d) + \frac{1}{n} \sum_{j=1}^{n} W_j(n) \binom{n}{j} p_d^j (1 - p_d)^{n-j}. \quad (9)$$

We also would like to make a few comments on the result of the Corollary 1. First of all, the lower bound (9) is tighter than the one proved in [3] ((1) with $p_i = p_e = 0$) which is the simplest analytical lower bound on the capacity of the deletion channel. The amount of improvement in (9) over the one in (1) is $\frac{1}{n} \sum_{j=1}^{n} W_j(n) \binom{n}{j} p_d^j (1 - p_d)^{n-j} - p_d$, which is guaranteed to be positive.

In [14], it is shown that

$$C_d = 1 + p_d \log(p_d) - A_1 p_d + O(p_d^{1.4}) \quad (10)$$

where $A_1 = \log(2e) - \sum_{l=1}^{\infty} 2^{-l-1} l \log(l)$. A similar result in [15] is provided, that is $C_d \leq 1 - (1 - O(p_d)) H_b(p_d)$, which shows that $1 - H_b(p_d)$ is a tight lower bound for small deletion probabilities. If we consider the new capacity lower bound in (9), and represent $(1 - p_d) \log(1 - p_d)$ by its Taylor series expansion, we can readily write

$$C_d \geq 1 + p_d \log(p_d) - (\log(2e) - W_1(n)) p_d + p_d^2 f(n, p_d)$$

where $f(n, p_d)$ is a polynomial function. On the other hand for $W_1(n)$, if we let n go to infinity, we have

$$\lim_{n \to \infty} W_1(n) = \lim_{n \to \infty} \left[\frac{1}{n} \sum_{l=1}^{n-1} 2^{-l-1} (n - l + 3) l \log(l) + \frac{\log(n)}{2^{n-1}} \right]$$
$$= \sum_{l=1}^{\infty} 2^{-l-1} l \log(l). \quad (11)$$

Therefore, we observe that the lower bound (9) captures the first-order term of the capacity expansion (10). This is an important result as the capacity expansions in [14] and [15] are asymptotic and do not lend themselves for a numerical calculation of the transmission rates for any nonzero value of the deletion probability.

We need the following two propositions in the proof of Lemma 1. In Proposition 1, we obtain the exact value of the output entropy in the deletion–substitution channel with i.u.d. input sequences, while Proposition 2 gives an upper bound on the conditional output entropy with i.u.d. bits transmitted through the deletion–substitution channel.

Proposition 1: For an i.i.d. deletion–substitution channel with i.u.d. input sequences of length n, we have

$$H(\boldsymbol{Y}') = n(1 - p_d) + H(\boldsymbol{T}) \quad (12)$$

where $\boldsymbol{Y}'$ denotes the output sequence of the deletion–substitution channel and $H(\boldsymbol{T})$ is as defined in (3).

Proof: By using the facts that all the elements of the set $\mathcal{Y}_{-j}^d$ are identically distributed, which are input into the BSC channel, and a fixed length i.u.d. input sequence into a BSC result in i.u.d. output sequences, all elements of the set $\mathcal{Y}'^d_{-j}$ are also identically distributed. Hence,

$$P(\boldsymbol{y}'(n - j)) = \frac{1}{2^{n-j}} \binom{n}{j} p_d^j (1 - p_d)^{n-j} \quad (13)$$

where $\binom{n}{j} p_d^j (1 - p_d)^{n-j}$ is the probability of exactly j deletions occurring in n use of the channel. Therefore, we obtain

$$H(\boldsymbol{Y}') = \sum_{\boldsymbol{y}'} -P(\boldsymbol{y}') \log(P(\boldsymbol{y}'))$$
$$= \sum_{j=0}^{n} \binom{n}{j} p_d^j (1 - p_d)^{n-j} \log \left(\frac{2^{n-j}}{\binom{n}{j} p_d^j (1 - p_d)^{n-j}} \right)$$
$$= n(1 - p_d) + H(\boldsymbol{T}) \quad (14)$$

which concludes the proof. ∎

Proposition 2: For a deletion–substitution channel with i.u.d. input sequences, the entropy of the output $\boldsymbol{Y}'$ conditioned on the input $\boldsymbol{X}$ of length n bits, is upper bounded by

$$H(\boldsymbol{Y}'|\boldsymbol{X}) \leq n H_b(p_d) - \sum_{j=1}^{n} W_j(n) \binom{n}{j} p_d^j (1 - p_d)^{n-j}$$
$$+ n(1 - p_d) H_b(p_e) \quad (15)$$

where $W_j(n)$ is given in (8).

Proof: To obtain the conditional output entropy, we need to compute the probability of all possible output sequences resulting from every possible input sequence $\boldsymbol{x}$, i.e., $P(\boldsymbol{Y}'|\boldsymbol{x})$. For a given $\boldsymbol{x} = (b; n_1, n_2, \ldots, n_k)$ and for a specific deletion pattern $D(n; K; j) = (j_1, \ldots, j_K)$ in which j_k denotes the number of deletions in the kth run, we can write

$$P\left(D(n; K; j) = (j_1, \ldots, j_K) \Big| \boldsymbol{x}(b; n_1, \ldots, n_K) \right)$$
$$= \binom{n_1}{j_1} \ldots \binom{n_K}{j_K} p_d^j (1 - p_d)^{n-j}. \quad (16)$$

Furthermore, for every $D(n; K; j)$, we can write

$$P\left(\boldsymbol{y}' \Big| D * \boldsymbol{x}(n; K) \right) = \begin{cases} p_e^s (1 - p_e)^{n-j-s} & \text{if } |\boldsymbol{y}'| = n - j, \\ 0 & \text{otherwise,} \end{cases} \quad (17)$$

where $s = d_H\left(\boldsymbol{y}'; D(n; K; j) * \boldsymbol{x}(n; K)\right)$, and $d_H = (\boldsymbol{a}; \boldsymbol{b})$ is the Hamming distance between two sequences $\boldsymbol{a}$ and $\boldsymbol{b}$. On the other hand, for every output sequence of length $n - j$, conditioned on a given input $\boldsymbol{x}(n; K)$, we have

$$P\left(\boldsymbol{y}'(n - j)\Big|\boldsymbol{x}(n; K)\right)$$
$$= \sum_{D \in \mathcal{D}_K^n(j)} P\left(\boldsymbol{y}'(n - j)\Big|D, \boldsymbol{x}(n; K)\right) P\left(D\Big|\boldsymbol{x}(n; K)\right).$$

However, there is a difficulty as two different possible deletion patterns, $D(n; K; j) = (j_1, \cdots, j_K)$ and $D'(n; K; j) = (j_1', \cdots, j_K')$, under the same substitution error pattern, i.e., the substitution errors occur at the same positions on $D(n; K; j) * x(n; K)$ and $D'(n; K; j) * x(n, K)$, may convert a given input sequence $\boldsymbol{x}(n; K)$ into the same output sequence, i.e., $D(n; K; j) * x(n; K) = D'(n; K; j) * x(n, K)$. This occurs when successive runs are completely deleted, for example, in transmitting $(1; 2, 1, 2, 3, 2) = 1101100011$, if the second, third, and fourth runs are completely deleted, by deleting one bit from the first run, $(1, 1, 2, 3, 0) * (1; 2, 1, 2, 3, 2) = (1; 1, 0, 0, 0, 2) = 111$, or from the last run, $(0, 1, 2, 3, 1) * (1; 2, 1, 2, 3, 2) = (1; 2, 0, 0, 0, 1) = 111$, the same output sequences are obtained. This difficulty can be addressed using

$$\sum_t -p_t \left(\log \sum_{t'} p_{t'}\right) \leq \sum_t -p_t \log(p_t) \qquad (18)$$

which is trivially valid for any set of probabilities $(p_1, \ldots, p_t, \ldots)$. Therefore, we can write

$$-P(\boldsymbol{y}'|\boldsymbol{x}) \log\left(P(\boldsymbol{y}'|\boldsymbol{x})\right)$$
$$= -\sum_{D \in \mathcal{D}_K^n(j)} P(\boldsymbol{y}'|D * \boldsymbol{x}) P(D|\boldsymbol{x}) \log\left(\sum_{D' \in \mathcal{D}_K^n(j)} P(\boldsymbol{y}'|D' * \boldsymbol{x}) P(D'|\boldsymbol{x})\right)$$
$$\leq -\sum_{D \in \mathcal{D}_K^n(j)} P(\boldsymbol{y}'|D * \boldsymbol{x}) P(D|\boldsymbol{x}) \log\left(P(\boldsymbol{y}'|D * \boldsymbol{x}) P(D|\boldsymbol{x})\right).$$
$$(19)$$

Hence, for a specific $\boldsymbol{x}(b; n; K^x) = (b; n_1^x, \ldots, n_{K^x}^x)$, we obtain (for more details see Appendix B)

$$H\left(\boldsymbol{Y}'\Big|\boldsymbol{x}(b; n; K^x)\right) \leq n H_b(p_d) + n(1 - p_d) H_b(p_e)$$
$$- \sum_{j=0}^n p_d^j (1 - p_d)^{n-j} \sum_{k=1}^{K^x} \sum_{j_k=0}^j \binom{n_k^x}{j_k}\binom{n - n_k^x}{j - j_K} \log\binom{n_k^x}{j_k}.$$

Therefore, by considering i.u.d. input sequences, we have

$$H(\boldsymbol{Y}'|\boldsymbol{X}) = \sum_{\boldsymbol{x} \in \mathcal{X}} \frac{1}{2^n} H(\boldsymbol{Y}'|\boldsymbol{x}) \leq n H_b(p_d) + n(1 - p_d) H_b(p_e)$$
$$- \sum_{j=0}^n \frac{p_d^j (1 - p_d)^{n-j}}{2^n} \sum_{\boldsymbol{x} \in \mathcal{X}} \sum_{k=1}^{K^x} \sum_{j_k=0}^j \binom{n_k^x}{j_k}\binom{n - n_k^x}{j - j_k} \log\binom{n_k^x}{j_k}.$$
$$(20)$$

On the other hand, we can write

$$\sum_{\boldsymbol{x} \in \mathcal{X}} \frac{1}{2^n} \sum_{k=1}^{K^x} \sum_{j_k=0}^j \binom{n_k^x}{j_k}\binom{n - n_k^x}{j - j_k} \log\binom{n_k^x}{j_k}$$
$$= \sum_{j'=0}^j \sum_{l=1}^n P_R(l, n)\binom{l}{j'}\binom{n-l}{j-j'} \log\binom{l}{j'} \qquad (21)$$

where $P_R(l, n)$ denotes the probability of having a run of length l in an input sequence of length n. It is obvious that $P_R(n, n) = \frac{2}{2^n}$. Due to the fact that, for $1 \leq l \leq n - 1$, there are $\binom{n-l-1}{K-2}$ possibilities of having a run of length l in a sequence with K runs, we can write

$$P_R(l, n) = \frac{2}{2^n} \sum_{K=2}^{n-l+1} \binom{n - l - 1}{K - 2} K = 2^{-l-1}(n-l+3). \quad (22)$$

Finally, by substituting (21) and (22) in (20), (15) results, completing the proof. ∎

We can now complete the proof of the main lemma of the section.

Proof of Lemma 1: In Theorem 1, we showed that for any input distribution and any transmission length, (3) results in a lower bound on the capacity of the channel with i.i.d. deletion errors. On the other hand, any lower bound on the information rate can also be used to derive a lower bound on the capacity. Due to the definition of the mutual information, (6), by obtaining the exact value of the output entropy (Proposition 1) and upper bounding the conditional output entropy (Proposition 2) the mutual information is lower bounded. Finally, by substituting (12) and (15) into (3), Lemma 1 is proved. □

At this point, we digress to point out that the result in the above lemma can also be obtained using a simpler approach as pointed out by one of the reviewers (details are given in Appendix A). That is, a lower bound on the deletion–substitution channel capacity can be provided in terms of the deletion-only channel capacity as (this is also a special case of a result in [16])

$$C_{ds} \geq C_d - (1 - p_d) H_b(p_e). \qquad (23)$$

Therefore, computing the mutual information rate of the deletion-only channel for i.u.d. input sequences and substituting it in the above inequality results in a lower bound on C_{ds}. It can be verified that the same procedure as in the proof of Lemma 1 gives

$$C_d \geq 1 - p_d - H_b(p_d) + \frac{1}{n} \sum_{j=1}^n W_j(n)\binom{n}{j} p_d^j (1 - p_d)^{n-j}$$

and substituting this into (23) concludes the proof of Lemma 1.

B. Deletion-AWGN Channel

In this section, a binary deletion channel in the presence of AWGN is considered, where the bits are transmitted using BPSK and the received signal contains AWGN in addition to
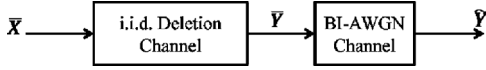
Fig. 2.   Deletion-AWGN channel as a cascade of an i.i.d. deletion channel and a BI-AWGN channel.

the deletion errors. As illustrated in Fig. 2, this channel can be considered as a cascade of two independent channels where the first channel is an i.i.d. deletion channel and the second one is a BI-AWGN channel. We use $\overline{X}$ to denote the input sequence to the first channel which is a BPSK modulated version of the binary input sequence X, i.e., $\bar{x}_i = 1 - 2x_i$, and $\overline{Y}$ to denote the output sequence of the first channel input to the second one. $\widetilde{Y}$ is the output sequence of the second channel that is the noisy version of $\overline{Y}$, i.e., $\widetilde{y}_i = \bar{y}_i + z_i$, in which z_is are i.i.d. Gaussian random variables with zero mean and a variance of σ^2, and $\widetilde{y}_i$ and $\bar{y}_i$ are the ith received and transmitted bits of the second channel, respectively. Therefore, for the probability density function of the ith channel output, we have

$$f_{\widetilde{y}_i}(\eta) = f_{\widetilde{y}_i}(\eta|\bar{y}_i = 1)P(\bar{y}_i = 1) + f_{\widetilde{y}_i}(\eta|\bar{y}_i = -1)P(\bar{y}_i = -1)$$
$$= \frac{1}{\sqrt{2\pi}\sigma}\left[P(\bar{y}_i = 1)e^{-\frac{(\eta-1)^2}{2\sigma^2}} + P(\bar{y}_i = -1)e^{-\frac{(\eta+1)^2}{2\sigma^2}}\right]. \tag{24}$$

In the following lemma, an achievable rate is provided over this channel.

Lemma 2: For any $n > 0$, the capacity of the deletion-AWGN channel with a deletion probability of p_d and a noise variance of σ^2 is lower bounded by

$$C_{d,AWGN} \geq 1 - p_d + \frac{1}{n}\sum_{j=1}^{n} W_j(n)\binom{n}{j}p_d^j(1-p_d)^{n-j}$$
$$- H_b(p_d) - (1-p_d)E\left[\log\left(1 + e^{\frac{-2z}{\sigma^2}}\right)\right] \tag{25}$$

where $W_j(n)$ is as given in (8), $E[.]$ is statistical expectation, and $z \sim \mathcal{N}(0, \sigma^2)$. $\square$

Before giving the proof of the above lemma, we provide several comments about the result. First, the desired lower bound in (25) is the only analytical lower bound on the capacity of the deletion-AWGN channel. In the current literature, there are only simulation-based lower bounds, e.g., [10], which employs Monte-Carlo simulation techniques. Furthermore, the procedure employed in [10] is only useful for deriving lower bounds for small values of deletion probability, e.g., $p_d \leq 0.1$, while the lower bound in (25) is useful for a much wider range. For $p_d = 0$, the lower bound in (25) is equal to $1 - E\left[\log(1 + e^{\frac{-2z}{\sigma^2}})\right]$ which is the capacity of the BI-AWGN channel [17, p. 362]. Finally, we note that the term in (25) which contains $E\left[\log(1 + e^{\frac{-2z}{\sigma^2}})\right]$ can be easily computed by numerical integration with an arbitrary accuracy (it involves only an 1-D integral).

We need the following two propositions in the proof of Lemma 2. In the following proposition, the exact value of the differential output entropy in the deletion-AWGN channel with i.u.d. input bits is calculated.

Proposition 3: For an i.i.d. deletion-AWGN channel with i.u.d. input sequences of length n, we have

$$h(\widetilde{Y}) = n(1 - p_d)\left(\log\left(2\sigma\sqrt{2\pi e}\right) - E\left[\log\left(1 + e^{-\frac{2z}{\sigma^2}}\right)\right]\right)$$
$$+ H(T) \tag{26}$$

where $h(.)$ denotes the differential entropy function, $\widetilde{Y}$ denotes the output of the deletion-AWGN channel, $z \sim \mathcal{N}(0, \sigma^2)$, and $H(T)$ is as defined in (3).

Proof: For the differential entropy of the output sequence, we can write

$$h(\widetilde{Y}) = h(\widetilde{Y}) + H(T|\widetilde{Y})$$
$$= h(\widetilde{Y}, T)$$
$$= h(\widetilde{Y}|T) + H(T) \tag{27}$$

where the first equality results by using the fact that by knowing the received sequence, the number of deletions is known and T is determined, i.e., $H(T|\widetilde{Y}) = 0$, and the last equality is obtained by using a different expansion of $h(\widetilde{Y}, T)$. On the other hand, we can write

$$h(\widetilde{Y}|T) = \sum_{j=0}^{n} h(\widetilde{Y}|T = j)P(T = j)$$
$$= \sum_{j=0}^{n} h(\widetilde{Y}|T = j)\binom{n}{j}p_d^j(1-p_d)^{n-j}. \tag{28}$$

Due to the fact that all the elements of the set $\overline{\mathcal{Y}}_{-j}^d$ are i.i.d., we have $P(\bar{y}(n-j)) = P(\bar{y}, T = j) = \frac{1}{2^{n-j}}\binom{n}{j}p_d^j(1-p_d)^{n-j}$. Therefore, we can write

$$P(\bar{y}|T = j) = \frac{P(\bar{y}, T = j)}{P(T = j)} = \frac{1}{2^{n-j}} \tag{29}$$

and as a result $P(\bar{y}_i = 1|T = j) = P(\bar{y}_i = -1|T = j) = \frac{1}{2}$ (for $1 \leq i \leq n - j$). By employing this result in (24), we have

$$f_{\widetilde{y}_i}(\eta) = \frac{1}{2\sqrt{2\pi}\sigma}\left[e^{-\frac{(\eta-1)^2}{2\sigma^2}} + e^{-\frac{(\eta+1)^2}{2\sigma^2}}\right] \tag{30}$$

where $f_{\widetilde{y}_i}(\eta)$ denotes the probability density function (PDF) of the continuous random variable $\widetilde{y}_i$. Noting also that the deletions happen independently and $\widetilde{y}_i$s are i.i.d., we can write

$$h(\widetilde{Y}|T = j) = (n - j)h(\widetilde{y}_i)$$
$$= (n - j)\int_{-\infty}^{\infty} -f_{\widetilde{y}_i}(\eta)\log\left(f_{\widetilde{y}_i}(\eta)\right)d\eta$$
$$= (n - j)\left(\log\left(2\sigma\sqrt{2\pi e}\right) - E\left[\log\left(1 + e^{-\frac{2z}{\sigma^2}}\right)\right]\right).$$

By substituting the above equation into (28), we obtain

$$h(\widetilde{Y}|T) = \sum_{j=0}^{n}(n - j)\binom{n}{j}p_d^j(1-p_d)^{n-j}$$
$$\times \left(\log(2\sigma\sqrt{2\pi e}) - E\left[\log\left(1 + e^{-\frac{2z}{\sigma^2}}\right)\right]\right)$$
$$= n(1 - p_d)\left(\log(2\sigma\sqrt{2\pi e}) - E\left[\log(1 + e^{-\frac{2z}{\sigma^2}})\right]\right) \tag{31}$$

and by using (31) and (27), (26) is obtained.   ∎

In the following proposition, we derive an upper bound on the differential entropy of the output conditioned on the input for deletion-AWGN channel.

Proposition 4: For a deletion-AWGN channel with i.u.d. input bits, the differential entropy of the output sequence $\widetilde{Y}$ conditioned on the input X of length n, is upper bounded by

$$h(\widetilde{Y}|X) \le nH_b(p_d) - \sum_{j=1}^{n} W_j(n) \binom{n}{j} p_d^j (1-p_d)^{n-j} \\ + n(1-p_d)\log(2\sigma\sqrt{2\pi e}) \tag{32}$$

where $W_j(n)$ is given in (8).

Proof: For the conditional differential entropy of the output sequence given the length n input X, we can write

$$h(\widetilde{Y}|X) = h(\widetilde{Y}|X) + H(T|\widetilde{Y},X) \\ = H(T) + h(\widetilde{Y}|T,X) \tag{33}$$

where the first equality follows since by knowing X and $\widetilde{Y}$, the number of deletions is known, i.e., $H(T|\widetilde{Y},X) = 0$. The second equality is obtained by using a different expansion of $h(\widetilde{Y},T|X)$ and also using the fact that the deletion process is independent of the input X, i.e., $H(T|X) = H(T)$. Furthermore, we have

$$h(\widetilde{Y}|T,X) = \sum_{j=0}^{n} h(\widetilde{Y}|X,T=j)P(T=j) \\ = \sum_{j=0}^{n} h(\widetilde{Y}|X,T=j)\binom{n}{j}p_d^j(1-p_d)^{n-j}.$$

To obtain $h(\widetilde{Y}|X,T=j)$, we need to compute $f_{\widetilde{y}|x,j}(\eta)$ for any given input sequence $x = (b; n_1, n_2, \ldots, n_K)$ and different values of j. As in the proof of Proposition 2, if we consider the outputs of the deletion channel resulting from different deletion patterns of length j from a given x, as if they are distinct and also use the result in (18), an upper bound on the differential output entropy conditioned on the input sequence X results. We relegate the details of this computation and completion of the proposition proof to Appendix C. ∎

We can now state the proof of the main lemma of the section.

Proof of Lemma 2: By substituting the exact value of the differential output entropy in (26), and the upper bound (32) on the differential output entropy conditioned on the input in (6), a lower bound on the mutual information rate of the deletion-AWGN channel is obtained, hence the lemma is proved. □

IV. LOWER BOUNDS ON THE CAPACITY OF RANDOM INSERTION CHANNELS

We now turn our attention to the random insertion channels and derive lower bounds on the capacity of random insertion channels by employing the approach proposed in Section II. We consider the Gallager model [3] for insertion channels in which every transmitted bit is independently replaced by two random bits with probability of p_i while neither the receiver nor the transmitter have any information about the position of the insertions. The following lemma provides the main result of this section.

Lemma 3: For any $n > 0$, the capacity of the random insertion channel C_i, is lower bounded by

$$C_i \ge (1-p_i)^n - H_b(p_i) + \left(S(n) - \frac{3n+1}{4n} + n\right)p_i(1-p_i)^{n-1}$$

$$+ \frac{\log\binom{n}{2}}{n}\left(1 - (1-p_i)^n - np_i(1-p_i)^{n-1} - p_i^n - np_i^{n-1}(1-p_i)\right)$$

$$+ p_i^{n-1}(1-p_i)\log(n) \tag{34}$$

where

$$S(n) = \frac{1}{4n}\sum_{l=1}^{n-1} 2^{-l}\left[(n+1-l)(l+2)\log(l+2) \\ + 2(l+1)\log(l+1)\right] + \frac{\log(n)}{2^{n+1}}. \quad \square$$

To the best of our knowledge, the only analytical lower bound on the capacity of the random insertion channel is derived in [3] (i.e., (1) for $p_d = p_e = 0$). Our result improves upon this result for small values of insertion probabilities as will be apparent with numerical examples.

Similar to the deletion–substitution channel case, we can write a simpler lower bound as

$$C_i \ge 1 - H_b(p_i) + \left(S(n) - \frac{3n+1}{4n}\right)p_i$$

$$- \frac{n-1}{2}\left(2S(n) - \frac{3n+1}{2n} + n - \log\binom{n}{2}\right)p_i^2$$

$$- \binom{n-1}{2}\left(\log\binom{n}{2} - S(n) - \frac{2n}{3} + \frac{3n+1}{4n}\right)p_i^3$$

$$- \binom{n-1}{3}\left(S(n) + n - \frac{3n+1}{4n}\right)p_i^4. \tag{35}$$

For instance, for $n = 10$, (35) evaluates to

$$C_i \ge 1 - H_b(p_i) + 1.1591p_i - 30.7184p_i^2 + 1.0502 \times 10^2 p_i^3 \\ - 1.3391 \times 10^3 p_i^4. \tag{36}$$

To prove the above lemma, we need the following two propositions. The output entropy of the random insertion channel with i.u.d. input sequences is calculated in the first one.

Proposition 5: For a random insertion channel with i.u.d. input sequences of length n, we have

$$H(Y) = n(1 + p_i) + H(T) \tag{37}$$

where Y denotes the output sequence and $H(T)$ is as defined in (3).

Proof: Similar to the proof of Proposition 1, we use the fact that

$$P(y(n+j)) = \frac{1}{2^{n+j}}\binom{n}{j}p_i^j(1-p_i)^{n-j}. \tag{38}$$

Therefore, by employing (38) in computing the output entropy, we obtain

$$
H(\boldsymbol{Y}) = -\sum_{j=0}^{n} \binom{n}{j} p_i^j (1-p_i)^{n-j} \log\left(\frac{\binom{n}{j} p_i^j (1-p_i)^{n-j}}{2^{n+j}}\right)
$$
$$
= n(1+p_i) + H(\boldsymbol{T}). \tag{39}
$$

■

In the following proposition, we present an upper bound on the conditional output entropy of the random insertion channel with i.u.d. input sequences for a given input of length n.

Proposition 6: For a random insertion channel with input and output sequences denoted by $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively, with i.u.d. input sequences of length n, we have

$$
H(\boldsymbol{Y}|\boldsymbol{X}) \le n(1+p_i) + n H_b(p_i) - n(1-p_i)^n
$$
$$
- \left(1 - (1-p_i)^n - np_i(1-p_i)^{n-1} - p_i^n - np_i^{n-1}(1-p_i)\right) \log\binom{n}{2}
$$
$$
- n\left(S(n) - \frac{3n+1}{4n} + n\right) p_i(1-p_i)^{n-1} - np_i^{n-1}(1-p_i)\log(n) \tag{40}
$$

where $S(n)$ is given in (34).

Proof: For the conditional output sequence distribution for a given input sequence, we can write

$$
p(\boldsymbol{y}|\boldsymbol{x}(b;n;K))
$$
$$
= \begin{cases}
(1-p_i)^n & \boldsymbol{y} = \boldsymbol{x}(b;n;K) \\
\frac{n_1+1}{4} p_i(1-p_i)^{n-1} & \boldsymbol{y} = (b; n_1+1, \ldots, n_K) \\
\frac{n_K+1}{4} p_i(1-p_i)^{n-1} & \boldsymbol{y} = (b; n_1, \ldots, n_K+1) \\
\frac{n_k+2}{4} p_i(1-p_i)^{n-1} & \boldsymbol{y} = (b; n_1, \ldots, n_k+1, \ldots, n_K) \\
& (1 < k < K) \\
\frac{1}{4} p_i(1-p_i)^{n-1} & \boldsymbol{y} = (b; n_1, \ldots, n'_{k,1}, 2, n'_{k,2}, \ldots, n_K) \\
\frac{2}{4} p_i(1-p_i)^{n-1} & \boldsymbol{y} = (b; n_1, \ldots, n''_{k,1}, 1, n''_{k,2}, \ldots, n_K) \\
\frac{1}{4} p_i(1-p_i)^{n-1} & \boldsymbol{y} = (\bar{b}; 1, n_1, \ldots, n_k, \ldots, n_K) \\
\frac{1}{4} p_i(1-p_i)^{n-1} & \boldsymbol{y} = (b; n_1, \ldots, n_k, \ldots, n_K, 1) \\
\epsilon_{y,x}^i & |\boldsymbol{y}| \ge n+2
\end{cases}
$$

where $n'_{k,1} + n'_{k,2} = n_k - 1$ ($n'_{k,1}, n'_{k,2} \ge 0$), $n''_{k,1} + n''_{k,2} = n_k$ ($n''_{k,1}, n''_{k,2} \ge 1$), and $\epsilon_{y,x}^i$ represents $p(\boldsymbol{y}|\boldsymbol{x}(b;n;K))$ for given $\boldsymbol{y}$ with $|\boldsymbol{y}| \ge n+2$. Furthermore, since there are n_k possibilities of $n'_{k,i} \ge 0$ having $n'_{k,1} + n'_{k,2} = n_k - 1$, and $n_k - 1$ possibilities of $n''_{k,i} \ge 1$ having $n''_{k,1} + n''_{k,2} = n_k$, we obtain

$$
H(\boldsymbol{Y}|\boldsymbol{x}(b;n;K^x)) = -(1-p_i)^n \log(1-p_i)^n
$$
$$
- p_i(1-p_i)^{n-1}\left(n\log(p_i(1-p_i)^{n-1}) - 1.5n - 0.5K^x\right)
$$
$$
- \frac{1}{4} p_i(1-p_i)^{n-1}\bigg((n_1^x+1)\log(n_1^x+1) + (n_{K^x}^x+1)\log(n_{K^x}^x+1)
$$
$$
+ \sum_{k=2}^{K^x-1} (n_k^x+2)\log(n_k^x+2)\bigg) + H_{\epsilon,i}(\boldsymbol{x}),
$$

where $H_{\epsilon,i}(\boldsymbol{x})$ is the term related to the outputs resulting from more than one insertion. Therefore, by considering i.u.d. input

sequences, since there are $2\binom{n-1}{K-1}$ input sequences of length n with K runs, we have

$$
H(\boldsymbol{Y}|\boldsymbol{X}) = -(1-p_i)^n \log(1-p_i)^n + H_{\epsilon,i}(\boldsymbol{X})
$$
$$
- np_i(1-p_i)^{n-1}\left(\log(p_i(1-p_i)^{n-1}) - \frac{7n+1}{4n} + S(n)\right) \tag{41}
$$

where $H_{\epsilon,i}(\boldsymbol{X}) = \sum_{\boldsymbol{x}\in\mathcal{X}} \frac{H_{\epsilon,i}(\boldsymbol{x})}{2^n}$ and

$$
S(n) = \frac{1}{2^{n+2}n} \sum_{x, K^x \ne 1} \bigg[(n_1^x+1)\log(n_1^x+1)
$$
$$
+ (n_{K^x}^x+1)\log(n_{K^x}^x+1) + \sum_{k=2}^{K^x-1}(n_k^x+2)\log(n_k^x+2)\bigg] + \frac{\log(n)}{2^{n+1}}
$$

which can be written as

$$
S(n) = \frac{\log(n)}{2^{n+1}} + \frac{1}{2^{n+2}n}\bigg[\sum_{\boldsymbol{x}}\sum_{k=1}^{K^x}(n_k^x+2)\log(n_k^x+2)
$$
$$
+ 2\sum_{x, K^x\ne 1}[(n_1^x+1)\log(n_1^x+1) - (n_1^x+2)\log(n_1^x+2)]\bigg]
$$
$$
= \frac{1}{4n}\sum_{l=1}^{n-1} 2^{-l}[(n+1-l)(l+2)\log(l+2) + 2(l+1)\log(l+1)]
$$
$$
+ \frac{\log(n)}{2^{n+1}}. \tag{42}
$$

Here we have used the same approach used in the proof of Proposition 2, and considered the fact that there are 2^{n-l} sequences of length n with $n_1 = l$ or $n_K = l$.

If we assume that all the possible outputs resulting from k insertions ($k \ge 2$) for a given $\boldsymbol{x}$ are equiprobable, since

$$
- \sum_{j=1}^{J} p_j \log(p_j) \le -\sum_{j=1}^{J} p_j \log\frac{\sum_{j'=1}^{J} p_{j'}}{J} \tag{43}
$$

we can upper bound $H_{\epsilon,i}(\boldsymbol{x})$. That is

$$
H_{\epsilon,i}(\boldsymbol{x}) = \sum_{k=2}^{n}\sum_{\boldsymbol{y}\in\mathcal{Y}^i(\boldsymbol{x}+k)} -Q(\boldsymbol{y}|\boldsymbol{x})\log\left(Q(\boldsymbol{y}|\boldsymbol{x})\right)
$$
$$
\le \sum_{k=2}^{n} -\epsilon_k \log\left(\frac{\epsilon_k}{|\mathcal{Y}^i(\boldsymbol{x}+k)|}\right) \le \sum_{k=2}^{n} -\epsilon_k \log\left(\frac{\epsilon_k}{2^{n+k}}\right)
$$

where $\epsilon_k = \sum_{\boldsymbol{y}\in(\boldsymbol{x},k)} Q(\boldsymbol{y}|\boldsymbol{x}) = \binom{n}{k} p_i^k(1-p_i)^{n-k}$ is the probability of k insertions in transmission of n bits, and the last inequality follows since $|\mathcal{Y}^i(\boldsymbol{x}+k)| \le 2^{n+k}$, where $|\mathcal{Y}^i(\boldsymbol{x}+k)|$ denotes the number of output sequences resulting from k insertions into a given input sequence $\boldsymbol{x}$. After some algebra, we arrive at

$$
H_{\epsilon,i}(\boldsymbol{X}) \le n(1+p_i) + n H_b(p_i) - n(1-p_i)^n
$$
$$
- (n+1)np_i(1-p_i)^{n-1} + (1-p_i)^n \log(1-p_i)^n
$$
$$
+ np_i(1-p_i)^{n-1}\log\left(p_i(1-p_i)^{n-1}\right)
$$
$$
- np_i^{n-1}(1-p_i)\log(n) - \left(1 - p_i^n - (1-p_i)^n\right)
$$
$$
- np_i(1-p_i)^{n-1} - np_i^{n-1}(1-p_i)\right)\log\binom{n}{2}.
$$

TABLE I
LOWER BOUNDS ON THE CAPACITY OF THE DELETION–SUBSTITUTION CHANNEL (IN THE LEFT-HAND SIDE TABLE "1-LOWER BOUND" IS REPORTED)

p_d	p_e	1−LB (2)	1−LB (7) $n=1000$	1−LB (7) $n=100$	p_d	p_e	LB (2)	LB (7) $n=1000$	LB (7) $n=100$
10^{-5}	10^{-5}	3.6104×10^{-4}	$\mathbf{3.5817 \times 10^{-4}}$	3.5834×10^{-4}	0.01	0.01	0.8392	**0.8419**	0.8418
10^{-5}	10^{-4}	1.6535×10^{-3}	$\mathbf{1.6506 \times 10^{-3}}$	1.6508×10^{-3}	0.01	0.03	0.7268	**0.7373**	0.7293
10^{-5}	10^{-3}	1.15881×10^{-2}	$\mathbf{1.15853 \times 10^{-2}}$	1.15854×10^{-2}	0.01	0.10	0.4549	**0.4576**	0.4575
10^{-4}	10^{-5}	1.6535×10^{-3}	$\mathbf{1.6248 \times 10^{-3}}$	1.6264×10^{-3}	0.05	0.01	0.6368	**0.6476**	0.6469
10^{-4}	10^{-4}	2.9459×10^{-3}	$\mathbf{2.9172 \times 10^{-3}}$	2.9188×10^{-3}	0.05	0.03	0.5289	**0.5397**	0.5390
10^{-4}	10^{-3}	1.2879×10^{-2}	$\mathbf{1.2850 \times 10^{-2}}$	1.2852×10^{-2}	0.05	0.10	0.2681	**0.2789**	0.2781
10^{-3}	10^{-5}	1.1588×10^{-2}	$\mathbf{1.1302 \times 10^{-2}}$	1.1319×10^{-2}	0.10	0.01	0.4583	**0.4729**	0.4716
10^{-3}	10^{-4}	1.2879×10^{-2}	$\mathbf{1.2593 \times 10^{-2}}$	1.261×10^{-2}	0.10	0.03	0.3561	**0.3707**	0.3693
10^{-3}	10^{-3}	2.2804×10^{-2}	$\mathbf{2.2518 \times 10^{-2}}$	2.2535×10^{-2}	0.10	0.10	0.1089	**0.1236**	0.1222

Finally, by substituting the above upper bound into (41), the upper bound (40) is obtained. ∎

Proof of Lemma 3: By substituting the exact value of the output entropy (37) and the upper bound on the conditional output entropy (40) of the random insertion channel with i.u.d. input sequences into (6), a lower bound on the achievable information rate is obtained, hence the lemma is proved. □

V. NUMERICAL EXAMPLES

We now present several examples of the lower bounds on the insertion/deletion channel capacity for different values of n and compare them with the existing ones in the literature.

A. Deletion–Substitution Channel

In Table I, we compare the lower bound (7) for $n = 100$ and $n = 1000$ with the one in [3]. We observe that the new bound improves the result of [3] for the entire range of p_d and p_e, and also as expected, by increasing n from 100 to 1000, a tighter lower bound for all values of p_d and p_e is obtained.

B. Deletion-AWGN Channel

We now compare the derived analytical lower bound on the capacity of the deletion-AWGN channel with the simulation-based bound of [10] which is the achievable information rate of the deletion-AWGN channel for i.u.d. input sequences obtained by Monte-Carlo simulations. As we observe in Fig. 3, the lower bound (25) is very close to the simulation results of [10] for small values of deletion probability but it does not improve them. This is not unexpected, because we further lower bounded the achievable information rate for i.u.d. input sequences while in [10], the achievable information rate for i.u.d. input sequences is obtained by Monte-Carlo simulations without any further lower bounding. On the other hand, new bound is provable, analytical, and very easy to compute while the result in [10] requires lengthy simulations. Furthermore, the procedure employed in [10] is only useful for deriving lower bounds for small values of deletion probability, e.g., $p_d \leq 0.1$, while the lower bound (25) holds for a much wider range.

C. Random Insertion Channel

We now numerically evaluate the lower bounds derived on the capacity of the random insertion channel. Similar to the previous cases, different values of n result in different lower bounds. In Table II and Fig. 4, we compare the lower bound in (34) with the lower bound due to Gallager [3] $C_i \geq 1 - H_b(p_i)$,
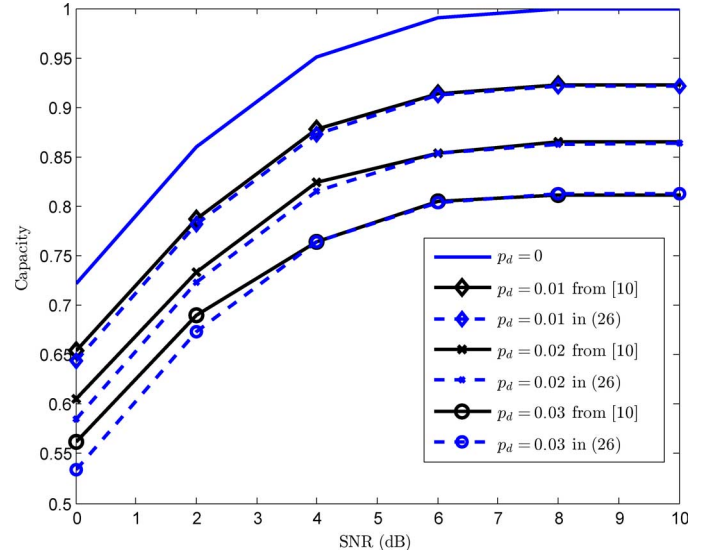

Fig. 3. Comparison between the lower bound (25) for $n = 1000$ with the lower bound in [10] versus SNR for different deletion probabilities.

where the reported values are obtained for the optimal value of n. We observe that for larger p_i, smaller values of n give the tightest lower bounds. This is not unexpected since in upper bounding $H(\boldsymbol{Y}|\boldsymbol{X})$, we computed the exact value of $p(\boldsymbol{y}|\boldsymbol{x})$ for at most one insertion, i.e., $|\boldsymbol{y}| = |\boldsymbol{x}|$ or $|\boldsymbol{y}| = |\boldsymbol{x}| + 1$, and upper bounded the part of the conditional entropy resulting from more than one insertion. Therefore, for a fixed p_i by increasing n, the probability of having more than one insertion increases and as a result the upper bound becomes loose. We also observe that the lower bound (34) improves upon the lower bound in [3] for $p_i < 0.25$, e.g., for $p_i = 0.1$, we achieve an improvement of 0.0392 bits/channel use.

VI. CONCLUSION

We have presented several analytical lower bounds on the capacity of the insertion/deletion channels by lower bounding the mutual information rate for i.u.d. input sequences. We have derived the first analytical lower bound on the capacity of the deletion-AWGN channel which for small values of deletion probability is very close to the existing simulation-based lower bounds. The lower bound presented on the capacity of the deletion–substitution channel improves the existing analytical lower bound for all values of deletion and substitution probabilities. For random insertion channel, the presented lower

TABLE II
LOWER BOUNDS ON THE CAPACITY OF THE RANDOM INSERTION CHANNEL (IN THE LEFT-HAND SIDE TABLE "1-LOWER BOUND" IS REPORTED)

p_i	$1-$LB from [3]	$1-$LB (34)	optimal value of n
10^{-6}	2.14×10^{-5}	$\mathbf{2.007 \times 10^{-5}}$	121
10^{-5}	1.81×10^{-4}	$\mathbf{1.68 \times 10^{-4}}$	57
10^{-4}	1.47×10^{-3}	$\mathbf{1.35 \times 10^{-3}}$	27
10^{-3}	1.14×10^{-2}	$\mathbf{1.02 \times 10^{-2}}$	13
10^{-2}	8.07×10^{-1}	$\mathbf{7.14 \times 10^{-2}}$	7

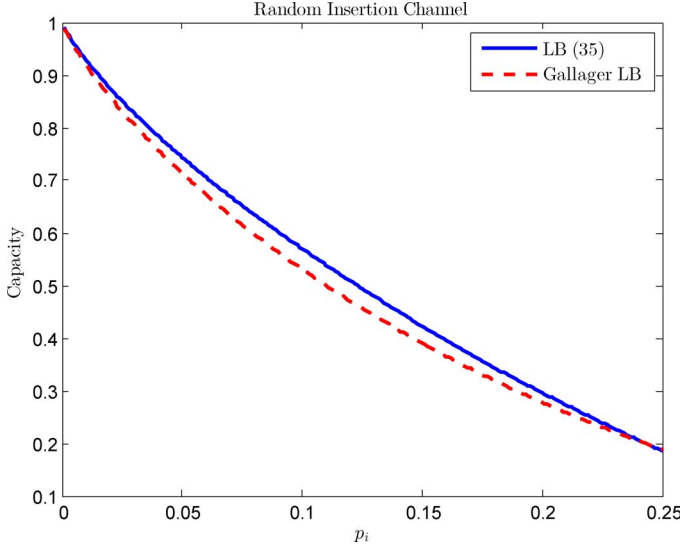| p_i	LB from [3]	LB (34)	optimal value of n
0.03	0.8056	**0.8276**	5
0.05	0.7136	**0.7442**	5
0.10	0.5310	**0.5702**	4
0.15	0.3901	**0.4230**	4
0.20	0.2781	**0.2962**	3
0.23	0.2220	**0.2283**	3
0.25	**0.1887**	0.1853	3

Fig. 4. Comparison of the lower bound (34) with lower bound presented in [3].

bound improve the existing ones for $p_i < 0.25$. For $p_e = 0$, the presented lower bound on the capacity of the deletion–substitution channel results into a lower bound on the capacity of the deletion-only channel which for small values of deletion probability, is very close to the tightest presented lower bounds, and is in agreement with the first order expansion of the channel capacity for $p_d \to 0$, while our result is a strict lower bound for the entire range of p_d.

APPENDIX A
DELETION–SUBSTITUTION CHANNEL CAPACITY IN TERMS OF THE DELETION CHANNEL CAPACITY

In this appendix, we relate the deletion–substitution and deletion-only channel capacities through an inequality (as pointed to us by one of the reviewers) which is a special case of a result obtained by the authors in [16]. This inequality can provide a tool to provide simpler proof for Lemma 1.

Claim 1: For any possible input distribution $P(\boldsymbol{X})$, we have

$$I(\boldsymbol{X};\boldsymbol{Y'}) \geq I(\boldsymbol{X};\boldsymbol{Y}) - n(1 - p_d)H_b(p_e). \qquad (44)$$

Proof: In Fig. 1, $\boldsymbol{X} \to \boldsymbol{Y} \to \boldsymbol{Y'}$ form a Markov chain. Let $\boldsymbol{F}$ be the "flipping" process of the BSC channel, consisting of $(1 - p_d + \delta)n$ bits drawn from i.i.d. Bernoulli(p_e), where

1 represents a flip, and 0 represents a location that is unaffected, and $\delta > 0$ is some constant we can choose later. Clearly, $\boldsymbol{Y'} = f(\boldsymbol{Y}; \boldsymbol{F})$ with high probability for the obvious function $f(.)$ which does $Y_i' = Y_i \oplus F_i$ for all bits in $\boldsymbol{Y}$. (There is a problematic event corresponding to more than $(1 - p_d + \delta)n$ bits passing through the deletion channel, but the probability of this event goes to 0 as $n \to \infty$. This event can be dealt with and we ignore it below, simply assuming $\boldsymbol{Y'} = f(\boldsymbol{Y}; \boldsymbol{F})$. Note that we also have $\boldsymbol{Y} = f(\boldsymbol{Y'}; \boldsymbol{F})$ at the same time).

Hence, for the mutual information $I(\boldsymbol{X};\boldsymbol{Y'})$, we have

$$I(\boldsymbol{X};\boldsymbol{Y'}) = H(\boldsymbol{X}) - H(\boldsymbol{X}|\boldsymbol{Y'})$$
$$= H(\boldsymbol{X}) - H(\boldsymbol{X}|\boldsymbol{Y'}, \boldsymbol{F}) - I(\boldsymbol{F};\boldsymbol{X}|\boldsymbol{Y'}).$$

Now, $H(\boldsymbol{X}|\boldsymbol{Y'}, \boldsymbol{F}) = H(\boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{Y'}, \boldsymbol{F}) = H(\boldsymbol{X}|\boldsymbol{Y})$ since $\boldsymbol{Y} = f(\boldsymbol{Y'}, \boldsymbol{F})$ and $\boldsymbol{X} \to \boldsymbol{Y} \to (\boldsymbol{F}, \boldsymbol{Y'})$ form a Markov chain. Further, $I(\boldsymbol{F};\boldsymbol{X}|\boldsymbol{Y'}) \leq H(\boldsymbol{F}|\boldsymbol{Y'}) = H(\boldsymbol{F}) = n(1 - p_d + \delta)H_b(p_e)$. It follows that

$$I(\boldsymbol{X};\boldsymbol{Y'}) \geq H(\boldsymbol{X}) - H(\boldsymbol{X}|\boldsymbol{Y}) - n(1 - p_d + \delta)H_b(p_e)$$
$$= I(\boldsymbol{X};\boldsymbol{Y}) - n(1 - p_d + \delta)H_b(p_e).$$

Since $\delta > 0$ is arbitrary, the result follows. ∎

Corollary 2: Let C_d and C_{ds} denote the deletion-only and deletion–substitution channel capacities, respectively, then

$$C_{ds} \geq C_d - (1 - p_d)H_b(p_e). \qquad (45)$$

Proof: Since (44) holds for any possible input distribution, it holds for capacity achieving input distribution for the deletion-only channel as well. Therefore, by dividing both sides by n and letting n go to infinity the proof follows. ∎

APPENDIX B
PART OF PROOF OF PROPOSITION 2

$$H\left(\boldsymbol{Y'}\Big|\boldsymbol{x}(b;n;K^x)\right)$$
$$= -\sum_{j=0}^{n}\sum_{\boldsymbol{y'}\in\mathcal{Y}_{-j}^d} P\left(\boldsymbol{y'}(n-j)|\boldsymbol{x}\right)\log\left(P\left(\boldsymbol{y'}(n-j)|\boldsymbol{x}\right)\right)$$
$$\leq -\sum_{j=0}^{n}\sum_{\boldsymbol{y'}\in\mathcal{Y}_{-j}^d}\sum_{D\in\mathcal{D}_K^n(j)} P(\boldsymbol{y'}|D*\boldsymbol{x})P(D|\boldsymbol{x})\log(P(\boldsymbol{y'}|D*\boldsymbol{x})P(D|\boldsymbol{x}))$$

where the inequality is obtained from the expression in (19). Furthermore, by employing the results from (16) and (17) and using the fact that there are $\binom{n-j}{s}$, distinct output sequences of length $n-j$ resulting from s substitution errors into a given input $\boldsymbol{x}$, i.e., $s = d_H\left(\boldsymbol{y}'(n-j); D(n;K;j) * \boldsymbol{x}(n;K)\right)$, we arrive at

$$
\begin{aligned}
&H\left(\boldsymbol{Y}'\Big|\boldsymbol{x}(b;n;K^x)\right) \\
&\leq -\sum_{j=0}^{n}\sum_{s=0}^{n-j}\binom{n-j}{s}\sum_{j_1+\cdots+j_K=j}p_e^s(1-p_e)^{n-j-s} \\
&\quad \times \binom{n_1^x}{j_1}\cdots\binom{n_K^x}{j_K}p_d^j(1-p_d)^{n-j} \\
&\quad \times \log\left(\binom{n_1^x}{j_1}\cdots\binom{n_K^x}{j_K}p_d^j(1-p_d)^{n-j}p_e^s(1-p_e)^{n-j-s}\right) \\
&= -\sum_{j=0}^{n}\sum_{j_1+\cdots+j_K=j}\binom{n_1^x}{j_1}\cdots\binom{n_K^x}{j_K}p_d^j(1-p_d)^{n-j} \\
&\quad \times \left[-(n-j)H_b(p_e)+\log\left(\binom{n_1^x}{j_1}\cdots\binom{n_K^x}{j_K}p_d^j(1-p_d)^{n-j}\right)\right] \\
&= nH_b(p_d)-\sum_{j=0}^{n}\sum_{j_1+\cdots+j_K=j}\binom{n_1^x}{j_1}\cdots\binom{n_K^x}{j_K}p_d^j(1-p_d)^{n-j} \\
&\quad \times \left[-n(1-p_d)H_b(p_e)+\log\left(\binom{n_1^x}{j_1}\cdots\binom{n_K^x}{j_K}\right)\right].
\end{aligned}
$$

Using the generalized Vandermonde's identity, that is,

$$
\sum_{j_1+\ldots+j_{K^x}=j}\binom{n_1^x}{j_1}\cdots\binom{n_{K^x}^x}{j_{K^x}}=\binom{n}{j}
$$

and the result

$$
\begin{aligned}
&\sum_{j_1+\ldots+j_{K^x}=j}\binom{n_1^x}{j_1}\cdots\binom{n_{K^x}^x}{j_{K^x}}\log\left(\binom{n_1^x}{j_1}\cdots\binom{n_{K^x}^x}{j_{K^x}}\right) \\
&=\sum_{j_1+\ldots+j_{K^x}=j}\binom{n_1^x}{j_1}\cdots\binom{n_{K^x}^x}{j_{K^x}}\sum_{k=1}^{K^x}\log\binom{n_k^x}{j_k} \\
&=\sum_{k=1}^{K^x}\sum_{j_k=0}^{j}\binom{n_k^x}{j_k}\binom{n-n_k^x}{j-j_k}\log\binom{n_k^x}{j_k}
\end{aligned}
$$

we obtain

$$
\begin{aligned}
&H\left(\boldsymbol{Y}'\Big|\boldsymbol{x}(b;n;K^x)\right)\leq nH_b(p_d)+n(1-p_d)H_b(p_e) \\
&-\sum_{j=0}^{n}p_d^j(1-p_d)^{n-j}\sum_{k=1}^{K^x}\sum_{j_k=0}^{j}\binom{n_k^x}{j_k}\binom{n-n_k^x}{j-j_k}\log\binom{n_k^x}{j_k}.
\end{aligned}
$$

APPENDIX C
PROOF OF PROPOSITION 4

For an i.i.d. deletion-AWGN channel, for a given $\boldsymbol{x}(b;n;K)$ and a fixed j, defining $\alpha(D,\boldsymbol{x})=1-2(D*\boldsymbol{x})$, i.e., $\alpha_i(D,\boldsymbol{x})\in\{1,-1\}$, yields

$$
\begin{aligned}
&f_{\widetilde{\boldsymbol{y}}}(\eta|\boldsymbol{x}(b;n;K),j) \\
&=\sum_{D\in\mathcal{D}_K^n(j)}f_{\widetilde{\boldsymbol{y}}}(\eta|\boldsymbol{x}(b;n;K),D)P(D|\boldsymbol{x}(b;n;K)) \\
&=\sum_{D\in\mathcal{D}_K^n(j)}f_{\widetilde{\boldsymbol{y}}}(\eta|\alpha(D,\boldsymbol{x}))P(D|\boldsymbol{x}(b;n;K)) \\
&=\sum_{D\in\mathcal{D}_K^n(j)}f_{\widetilde{y}_1\ldots\widetilde{y}_{n-j}}(\eta_1\ldots\eta_{n-j}|\alpha_1\ldots\alpha_{n-j})P(D|\boldsymbol{x}(b;n;K)) \\
&=\sum_{D\in\mathcal{D}_K^n(j)}f_{\widetilde{y}_1}(\eta_1|\alpha_1)\ldots f_{\widetilde{y}_{n-j}}(\eta_{n-j}|\alpha_{n-j})P(D|\boldsymbol{x}(b;n;K))
\end{aligned}
$$

where the last equality follows the fact that the noise samples $\boldsymbol{z}_i$ are independent and $\alpha_i(D,\boldsymbol{x})$ are also independent. By employing

$$
f_{\widetilde{y}_i}(\eta_i|\alpha_i(D,\boldsymbol{x}))=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(\eta_i-\alpha_i(D,\boldsymbol{x}))^2}{2\sigma^2}\right),
$$

and $P\left(D(n;K;j)\Big|\boldsymbol{x}(b;n;K),j\right)=\frac{\binom{n_1}{j_1}\ldots\binom{n_K}{j_K}}{\binom{n}{j}}$, we can write

$$
\begin{aligned}
&f_{\widetilde{\boldsymbol{y}}}(\eta|\boldsymbol{x}(b;n;K),j) \\
&=\frac{1}{(\sqrt{2\pi}\sigma)^{n-j}}\sum_{D\in\mathcal{D}_K^n(j)}\prod_{i=1}^{n-j}e^{\frac{-(\eta_i-\alpha_i(D,\boldsymbol{x}))^2}{2\sigma^2}}P(D|\boldsymbol{x}(b;n;K),j) \\
&=\frac{1}{(\sqrt{2\pi}\sigma)^{n-j}}\sum_{j_1+\ldots+j_K=j}\frac{\binom{n_1}{j_1}\ldots\binom{n_K}{j_K}}{\binom{n}{j}}\prod_{i=1}^{n-j}e^{\frac{-(\eta_i-\alpha_i(D,\boldsymbol{x}))^2}{2\sigma^2}}.
\end{aligned}
$$

Therefore, by defining

$$
A(j_1,\ldots,j_K)=\frac{\binom{n_1}{j_1}\ldots\binom{n_K}{j_K}}{\binom{n}{j}}\prod_{i=1}^{n-j}e^{\frac{-(\eta_i-\alpha_i(D,\boldsymbol{x}))^2}{2\sigma^2}}
$$

we obtain

$$
\begin{aligned}
&h(\widetilde{\boldsymbol{Y}}|\boldsymbol{x},j) \\
&=-\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\frac{1}{(\sqrt{2\pi}\sigma)^{n-j}}\sum_{j_1+\ldots+j_K=j}A(j_1,\ldots,j_K) \\
&\quad \times\log\left(\frac{1}{(\sqrt{2\pi}\sigma)^{n-j}}\sum_{j_1'+\ldots+j_K'=j}A(j_1',\ldots,j_K')\right)d\eta_1\ldots d\eta_{n-j} \\
&=-\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\frac{1}{(\sqrt{2\pi}\sigma)^{n-j}}\sum_{j_1+\ldots+j_K=j}A(j_1,\ldots,j_K) \\
&\quad \times\left[\log\left(\sum_{j_1'+\ldots+j_K'=j}A(j_1',\ldots,j_K')\right)\right]d\eta_1\ldots d\eta_{n-j} \\
&\quad +(n-j)\log(\sqrt{2\pi}\sigma)
\end{aligned}
$$

where we used the result of the generalized Vandermonde's identity and also the fact that $\int_{-\infty}^{\infty} f_{\widetilde{y}_i}(\eta_i|\bar{y}_i)d\eta_i = 1$. By using the inequality

$$\sum_{j'_1+\ldots+j'_K=j} A(j'_1,\ldots,j'_K) \geq A(j_1,\ldots,j_K),$$

which holds for every $j_1 + \ldots + j_K = j$, we can write

$$h(\widetilde{\boldsymbol{Y}}|\boldsymbol{x},j) \leq (n-j)\log(\sqrt{2\pi}\sigma)$$
$$- \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(\sqrt{2\pi}\sigma)^{n-j}} \sum_{j_1+\ldots+j_K=j} A(j_1,\ldots,j_K)$$
$$\times \log\left(A(j_1,\ldots,j_K)\right) d\eta_1 \ldots d\eta_{n-j}$$
$$= (n-j)\log(\sqrt{2\pi e}\sigma) + \log\binom{n}{j}$$
$$- \sum_{j_1+\ldots+j_K=j} \frac{\binom{n_1}{j_1}\ldots\binom{n_K}{j_K}}{\binom{n}{j}} \log\left(\binom{n_1}{j_1}\ldots\binom{n_K}{j_K}\right).$$

By considering i.u.d. input sequences, we have

$$h(\widetilde{\boldsymbol{Y}}|\boldsymbol{X},T) = \sum_{j=0}^{n} \binom{n}{j} p_d^j (1-p_d)^{n-j} \sum_{\boldsymbol{x}\in\mathcal{X}} \frac{1}{2^n} h(\widetilde{\boldsymbol{Y}}|\boldsymbol{x},j)$$
$$\leq n(1-p_d)\log(\sqrt{2\pi e}\sigma)$$
$$+ \sum_{j=0}^{n} \binom{n}{j} p_d^j (1-p_d)^{n-j} \left[\log\binom{n}{j} - W_j(n)\right]$$
$$(46)$$

where $W_j(n)$ is given in (8), and the result is obtained by following the same steps as in the computation leading to (20). Therefore, by substituting (46) into (33), (32) is obtained which concludes the proof.

ACKNOWLEDGMENT

We would like to thank the editor and the reviewers for detailed comments on the manuscript. In particular, we would like to acknowledge that the simpler proof of Lemma 1 given in Appendix A is due to one of the reviewers.

REFERENCES

[1] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Probl. Inf. Transmiss.*, vol. 3, no. 4, pp. 11–26, 1967.
[2] R. L. Dobrushin, "General formulation of Shannon's main theorem on information theory," *Amer. Math. Soc. Trans.*, vol. 33, pp. 323–438, 1963.
[3] R. Gallager, Sequential decoding for binary channels with noise and synchronization errors MIT Lincoln Lab., Group Report, Tech. Rep. 2502, 1961.
[4] S. Diggavi and M. Grossglauser, "On transmission over deletion channels," in *Proc. Annu. Allerton Conf. Commun. Control Comput.*, 2001, vol. 39, pp. 573–582.
[5] S. Diggavi and M. Grossglauser, "On information transmission over a finite buffer channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1226–1237, Mar. 2006.
[6] E. Drinea and M. Mitzenmacher, "On lower bounds for the capacity of deletion channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4648–4657, Oct. 2006.
[7] E. Drinea and M. Mitzenmacher, "Improved lower bounds for i.i.d. deletion and insertion channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2693–2714, Aug. 2007.
[8] A. Kirsch and E. Drinea, "Directly lower bounding the information capacity for channels with i.i.d. deletions and duplications," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 86–102, Jan. 2010.
[9] A. Kavcic and R. H. Motwani, "Insertion/deletion channels: Reduced-state lower bounds on channel capacities," in *Proc. IEEE Int. Symp. Inf. Theory*, 2004, p. 229.
[10] J. Hu, T. M. Duman, M. F. Erden, and A. Kavcic, "Achievable information rates for channels with insertions, deletions and intersymbol interference with i.i.d. inputs," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1102–1111, Apr. 2010.
[11] D. Fertonani and T. M. Duman, "Novel bounds on the capacity of the binary deletion channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2753–2765, Jun. 2010.
[12] S. Diggavi, M. Mitzenmacher, and H. Pfister, "Capacity upper bounds for deletion channels," in *Proc. Int. Symp. Inf. Theory*, 2007, pp. 1716–1720.
[13] D. Fertonani, T. M. Duman, and M. F. Erden, "Bounds on the capacity of channels with insertions, deletions and substitutions," *IEEE Trans. Commun.*, vol. 59, no. 1, pp. 2–6, 2011.
[14] Y. Kanoria and A. Montanari, "On the deletion channel with small deletion probability," in *Proc. Int. Symp. Inf. Theory*, Jun. 2010, pp. 1002–1006.
[15] A. Kalai, M. Mitzenmacher, and M. Sudan, "Tight asymptotic bounds for the deletion channel with small deletion probabilities," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 997–1001.
[16] M. Rahmati and T. M. Duman, "On the capacity of binary input symmetric q-ary output channels with synchronization errors," in *Proc. Int. Symp. Inf. Theory*, Jul. 2012, pp. 691–695.
[17] J. G. Proakis, *Digital Communications*, 5th ed. New York, NY: McGraw-Hill, 2007.

Mojtaba Rahmati (S'11) received his B.S. degree in electrical engineering in 2007 from University of Tehran, Iran, his M.S. degree in telecommunication systems in 2009 from Sharif University of Technology, Tehran, Iran, and his Ph.D. in electrical engineering in 2013 from Arizona State University, Tempe, AZ. He is currently a research assistant at Arizona State University. His research interests include information theory, digital communications and digital signal processing.

Tolga M. Duman (S'95–M'98–SM'03–F'11) is a Professor of Electrical and Electronics Engineering Department of Bilkent University in Turkey, and is on leave from the School of ECEE at Arizona State University. He received the B.S. degree from Bilkent University in Turkey in 1993, M.S. and Ph.D. degrees from Northeastern University, Boston, in 1995 and 1998, respectively, all in electrical engineering. Prior to joining Bilkent University in September 2012, he has been with the Electrical Engineering Department of Arizona State University first as an Assistant Professor (1998–2004), then as an Associate Professor (2004–2008), and starting August 2008 as a Professor. Dr. Duman's current research interests are in systems, with particular focus on communication and signal processing, including wireless and mobile communications, coding/modulation, coding for wireless communications, data storage systems and underwater acoustic communications.

Dr. Duman is a Fellow of IEEE, a recipient of the National Science Foundation CAREER Award and IEEE Third Millennium medal. His publications include a book on MIMO Communications (by Wiley in 2007), over fifty journal papers and over one hundred conference papers. He served as an editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2003–08), IEEE TRANSACTIONS ON COMMUNICATIONS (2007–2012) and *IEEE Online Journal of Surveys and Tutorials* (2002–07). He is currently the coding and communication theory area editor for IEEE TRANSACTIONS ON COMMUNICATIONS (2011–present) and an editor for Elsevier Physical Communications Journal (2010–present).